Study on Geometrically Constrained IVA with Auxiliary Function Approach and VCD for In-Car Communication

Kana Goto, Li Li, Riki Takahashi, Shoji Makino and Takeshi Yamada
University of Tsukuba, Japan

***Abstract*—In this paper, we apply a geometrically constrained independent vector analysis (GCIVA) method to an in-car speech enhancement system and confirm its effectiveness in realistic environments. Specifically, we employ GCIVA with the auxiliary function approach and vectorwise coordinate descent (GCAV-IVA) to enhance the target speech in in-car communication, where multiple co-occurring speeches are recorded with a triangle microphone array. GCAV-IVA is a recently proposed geometrically constrained blind source separation method, that has been shown to be powerful in directional speech enhancement with a limited number of microphones. Moreover, it is noteworthy for its fast convergence, low computational cost, and non requirement of step-size tuning, which makes it suitable for practical applications. However, the experiments using this method were only conducted using simulated impulse responses (IRs). In this study we investigates GCAV-IVA using measured in-car IRs to simulate more realistic environments. Moreover, we apply GCAV-IVA in a data-adaptive manner. The experimental results revealed that GCAV-IVA significantly outperformed conventional beamforming methods in terms of signal-to-distortion ratio.**

I. Introduction

When capturing a speech using a distant microphone, the quality of the speech degrades significantly owing to the presence of noise and interference, thereby giving rise to a need for speech enhancement applications in noisy environments, such as in a car. In a speedily moving car, the noise from the engine and wind increases the difficulty of human communication. On the other hand, multiple co-occurring speeches reduce the speech recognition accuracy of navigation systems. To improve the quality of in-car communication, speech enhancement methods to denoise a captured speech have been widely studied [1], [2], [3], [4].

For stationary or slowly varying additive noise, such as road and wind noise, spectral subtraction is a simple and efficient method [5]. For nonstationary noise, such as interference from other passengers, beamforming is a widely used approach [6], [7]. Considering the property that relative positions from speakers to microphones in a car are usually constant, constructing a beamformer that steers a beam to the target direction to enhance the signal or a null to the interference direction to suppress the signal is a reasonable choice. However, this approach usually requires a large number of microphones or training samples of both speech and noise to achieve appreciable enhancement performance.

Other promising directional speech enhancement methods include geometrically constrained blind source separation (BSS) [8], [9], which exploits spatial information to guide the separation matrices to obtain a signal from a desired direction. Since geometrically constrained BSS usually separates signals using a spatial null, which is estimated on the basis of the statistical independence of source signals, it can work with a small number of microphones without any training samples. Geometrically constrained independent vector analysis (GCIVA) [10], [11] is one such method, which combines the optimization problem of IVA [12], [13] with beamforming-based geometric constraints derived from the prior spatial information of source signals and the sensor geometry. In [11], a parameter estimated algorithm called GCAV-IVA has been derived the basis of the auxiliary function approach [14] and vectorwise coordinate descent (VCD) [15], [16], which is noteworthy for its fast convergence, low computational cost, and non requirement of step-size parameter. These characteristics make GCAV-IVA suitable for practical applications. Furthermore, owing to the well-designed geometric constraints, GCAV-IVA can reduce the negative impact of block permutation between the low- and high-frequency bands in the auxiliary function-based IVA (AuxIVA) [17], and subsequently achieve better speech enhancement performance. The original paper [11] has experimentally confirmed the effectiveness of GCAV-IVA in simulated situations. We conducted experiments in more realistic environments. In particular, we investigated an in-car speech enhancement system using GCAV-IVA with a triangle microphone array, where we generated test data using in-car impulse responses (IRs) measured under several conditions. Besides applying GCAV-IVA, which uses a given direction to conduct the constraints, we apply it in a data-adaptive manner, where the spatial information is learned from training data samples. We compare GCAV-IVA with AuxIVA and conventional beamforming methods, including both fixed and adaptive beamformers.

II. Geometrically Constrained Independent Vector Analysis

A. Formulation

Let us consider a determined situation where I sources are observed by J microphones. Let $x_i(\omega, t)$ and $y_j(\omega, t)$ denote the short-time Fourier transform (STFT) coefficients of the signals observed at the i-th microphone and the j-th estimated sources, respectively. Here, ω and t are the frequency and time

978-988-14768-8-3/20/$31.00 ©2020 APSIPA

indices, respectively. We respectively denote the frequency-wise vector representation of the observations and the estimated sources by

$$\boldsymbol{x}(\omega,t) = [x_1(\omega,t),\dots,x_I(\omega,t)]^{\mathsf{T}} \in \mathbb{C}^I, \tag{1}$$

$$\boldsymbol{y}(\omega,t) = [y_1(\omega,t),\dots,y_J(\omega,t)]^{\mathsf{T}} \in \mathbb{C}^J, \tag{2}$$

where $J = I$ and $(\cdot)^{\mathsf{T}}$ denotes the transpose. When the STFT window length is sufficiently longer than the impulse responses between sources and microphones, the relationship between the observations and the estimated sources can be expressed with the time-invariant instantaneous mixture model as

$$\boldsymbol{y}(\omega,t) = \boldsymbol{W}(\omega)\boldsymbol{x}(\omega,t), \tag{3}$$

where $\boldsymbol{W}(\omega) = [\boldsymbol{w}_{\boldsymbol{1}}(\omega),\dots,\boldsymbol{w}_{\boldsymbol{I}}(\omega)]^{\mathsf{H}}$ is an $I \times I$ separation matrix and $(\cdot)^{\mathsf{H}}$ denotes the Hermitian transpose.

IVA assumes that sources follow a multivariate distribution and thus dependencies over frequency components can be exploited to avoid the permutation problem. The separation matrices $\mathcal{W} = \{\boldsymbol{W}(\omega)\}_\omega$ are estimated by minimizing the following objective function

$$J_{\text{IVA}}(\mathcal{W}) = \sum_{j=1}^{J} \mathbb{E}[G(\boldsymbol{y}_j(t))] - \sum_{\omega=1}^{\Omega} \log|\det \boldsymbol{W}(\omega)|, \tag{4}$$

where Ω denotes the number of frequency bins. $\mathbb{E}[\cdot]$ denotes the expectation operator and $\boldsymbol{y}_j(t)$ is the source-wise vector representation defined as

$$\boldsymbol{y}_j(t) = [y_j(1,t),\dots,y_j(\Omega,t)]^{\mathsf{T}} \in \mathbb{C}^\Omega. \tag{5}$$

Here, $G(\boldsymbol{y}_j(t))$ is the contrast function having the relationship $G(\boldsymbol{y}_j(t)) = -\log p(\boldsymbol{y}_j(t))$, where $p(\boldsymbol{y}_j(t))$ represents a multivariate probability density function of the j-th source. One typical choice of the contrast function is to use a spherical multivariate distribution [12], [13], [17], which is expressed as

$$G(\boldsymbol{y}_j(t)) = G_R(r_j(t)), \tag{6}$$

$$r_j(t) = ||\boldsymbol{y}_j(t)||_2 = \sqrt{\sum_\omega |y_j(\omega,t)|^2}. \tag{7}$$

Here, $||\cdot||_2$ denotes the L_2 norm of a vector.

Now, let us consider a geometric constraint [8] that restricts the far-field response of the j-th separation filter estimated by IVA in the direction θ, which is described as

$$J_c(\mathcal{W}) = \sum_{j=1}^{J} \lambda_j \sum_{\omega=1}^{\Omega} |\boldsymbol{w}_j^{\mathsf{H}}(\omega)\boldsymbol{d}_j(\omega,\theta) - c_j|^2. \tag{8}$$

Here, $\boldsymbol{d}_j(\omega,\theta)$ is the steering vector pointing to the direction θ, c_j is the nonnegative-valued constraint, and $\lambda_j \geq 0$ is a parameter that weighs the importance of the constraint. This concept is used in the linearly constrained minimum variance (LCMV) beamformer [18]. Note that (8) with $c_j = 1$ forces the spatial filter to form a conventional delay-and-sum (DS) beamformer steering in the direction θ to preserve the target source whereas a small value of c_j essentially creates a spatial null towards the target direction θ, aiming at suppressing the target source and preserving all other sources. The null constraint on the target direction can also serve as a blocking matrix (BM) [19], so that the corresponding channel can produce a good estimate of interference and noise. Such an estimate would have the potential benefit of better handling under/overdetermined cases than conventional BSS methods, making it viable for practical applications. The objective function of the GCIVA is summarized as

$$J(\mathcal{W}) = J_{\text{IVA}}(\mathcal{W}) + J_c(\mathcal{W}). \tag{9}$$

B. Inference algorithm with auxiliary function approach

To explore the benefits of fast convergence and non requirement of a step-size parameter, the inference algorithm of GCAV-IVA is derived using the auxiliary function approach [20]. In this approach, the auxiliary function $J^+(\mathcal{W},\mathcal{V})$ is designed in such a way that $J(\mathcal{W}) = \min_{\mathcal{V}} J^+(\mathcal{W},\mathcal{V})$ is satisfied. Then, instead of directly optimizing the original objective function (9), which is difficult to analytically solve, the auxiliary function $J^+(\mathcal{W},\mathcal{V})$ is alternately minimized in terms of $\mathcal{W}$ and $\mathcal{V}$.

Since the geometric constraints are linear, we can simply obtain the auxiliary function that upper-bounds (9) by combining the original AuxIVA's auxiliary function [17] with the linear constraints:

$$\begin{aligned} J^+(\mathcal{W},\mathcal{V}) \overset{c}{=} \sum_{j=1}^{J}\sum_{\omega=1}^{\Omega} \Big\{ \frac{1}{2} \sum_j \boldsymbol{w}_j^{\mathsf{H}}(\omega)\boldsymbol{V}_j(\omega)\boldsymbol{w}_j(\omega) \\ - \log|\det \boldsymbol{W}(\omega)| \Big\} + J_c(\mathcal{W}), \end{aligned} \tag{10}$$

where $\boldsymbol{V}_j(\omega)$ is the weighted covariances expressed as

$$\boldsymbol{V}_j(\omega) = \mathbb{E}\Big[\frac{G_R^{'}(r_j(t))}{r_j(t)}\boldsymbol{x}(\omega)\boldsymbol{x}^{\mathsf{H}}(\omega)\Big] \tag{11}$$

and $=^c$ denotes equality up to constant terms. Here, $(\cdot)'$ denotes the derivative operator. When using the source model $G_R(r_j(t)) = r_j(t)$, $\boldsymbol{V}_j(\omega)$ can be expressed as $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}/r_j(t)]$.

The update rule for $\mathcal{V}$ is obtained straightforwardly by applying (7) into (11), whereas the update rule for $\mathcal{W}$ is derived by embracing the idea adopted in VCD [15], [16] that arranges the term $\log|\det \boldsymbol{W}|$ with the property of cofactor expansion. With the indices ω and θ omitted to simplicity the notation, the derived update rules are summarized as follows:

$$\boldsymbol{u}_j = \boldsymbol{D}_j^{-1}\boldsymbol{W}^{-1}\boldsymbol{e}_j, \tag{12}$$

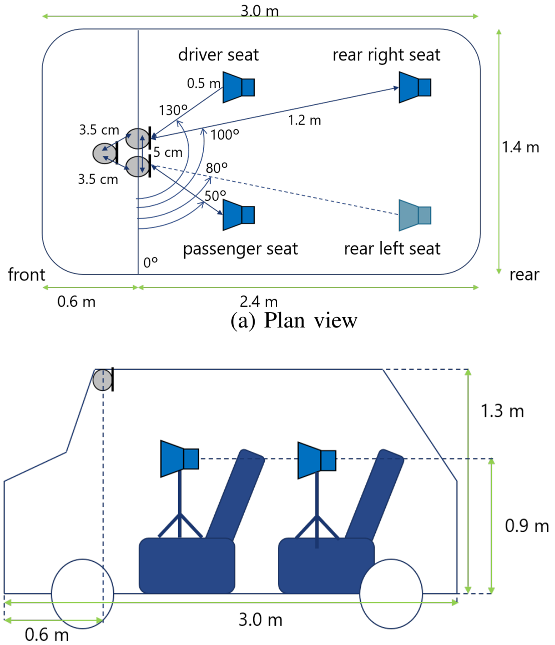
$$\hat{\boldsymbol{u}}_j = \lambda_j c_j \boldsymbol{D}_j^{-1}\boldsymbol{d}_j, \tag{13}$$

$$h_j = \boldsymbol{u}_j^{\mathsf{H}}\boldsymbol{D}_j\boldsymbol{u}_j, \tag{14}$$

$$\hat{h}_j = \boldsymbol{u}_j^{\mathsf{H}}\boldsymbol{D}_j\hat{\boldsymbol{u}}_j, \tag{15}$$

$$\boldsymbol{w}_j = \begin{cases} \frac{1}{\sqrt{h_j}}\boldsymbol{u}_j + \hat{\boldsymbol{u}}_j \ \text{(if } \hat{h}_j = 0), \\ \frac{\hat{h}_j}{2h_j}\Big[-1 + \sqrt{1 + \frac{4h_j}{|\hat{h}_j|^2}}\Big]\boldsymbol{u}_j + \hat{\boldsymbol{u}}_j \ \text{(o.w.)}. \end{cases} \tag{16}$$

Here, $\boldsymbol{D}_j = \boldsymbol{V}_j + \lambda_j\boldsymbol{d}_j\boldsymbol{d}_j^{\mathsf{H}}$ and $\boldsymbol{e}_j$ is the j-th column of the $I \times I$ identity matrix. Note that these update rules are equivalent to those employed in AuxIVA when $\lambda_j = 0$. The details of the derivation are available in [11] and [15].



(a) Plan view

(b) Side view

Fig. 1: Layout of sound sources and microphones

C. Fixed and data-adaptive constraints

Similary to the minimum variance distortionless response (MVDR) beamformer [6], [21], there are multiple ways to obtain the steering vector $\boldsymbol{d}_j(\omega, \theta)$. In this work, we apply GCAV-IVA with constraints conducted using two different steering vector estimation methods. The first one is based on the direction of arrival (DOA) under the plane wave propagation assumption with the prior knowledge of the microphone array geometry, which we refer to as fixed constraints. The second one is based on the eigenvalue decomposition of the observed speech covariance matrix $\boldsymbol{R}_s$. The estimated steering vector is given as the principal eigenvector of the speech covariance matrix as

$$\boldsymbol{d}(\omega) = \mathrm{PE}\{\boldsymbol{R}_s\}, \tag{17}$$

where $\mathrm{PE}\{\cdot\}$ is the operation to extract the principal eigenvector of a matrix. We refer to the second method as GCAV-IVA with data-adaptive constraints since the steering vector is estimated using the training data containing the speech active period. Adaptive beamformers need training samples containing both the target active period and interferer active period, whereas GCAV-IVA needs only the target active period.

III. Experiments

To evaluate the effectiveness of GCAV-IVA for in-car communication, we conducted speech enhancement experiments using the measured in-car IRs.

A. Datasets

We used speech signals from 10 speakers (6 males and 4 females) extracted from the ATR Japanese Speech Database [22], which consists of 503 phoneme-balanced sentences spoken by each speaker. The audio files are about 20 seconds long. We used the triangle microphone array set at the map lamp to record time-stretched pulse (TSP) signals played by loudspeakers placed at the driver, passenger, and rear left/right seats to measure the IRs in a car. The loudspeakers were placed at the center, left-of-center, or right-of-center of seats. We measured IRs under two conditions where all the car windows were open and closed. The details of the car and microphone array are shown in Fig. 1. The reverberation time (T_{60}) was 58 ms. This corresponds to both of open and closed windows.

TABLE I: Experimental conditions

Number of microphones	3
Number of sources	3
Reverberation time T_{60}	58 ms
DOA of the target	130°
DOAs of interferers	50°, 80°/ 50°, 100°
STFT flame length	1024 samples
STFT shift	256 samples
Training data length (target/interference)	5 sec/5 sec
Test data length	5 sec

TABLE II: Hyperparameters of GCAV-IVA

System #	c [tgt i_1 i_2]	λ [tgt i_1 i_2]
(1) fixed constraints	[1 0.2 0.2]	[0.1 0.1 0.1]
(2) adaptive constraints	[1 0.2 0.2]	[1 1 1]

We convolved the measured IRs with randomly selected speeches and added them together to generate mixture signals of three speakers. The target and interferers were respectively assumed to be the driver, and passengers sitting in the passenger seat and rear left or right seat. Thus, the DOA of the target was about 130°, whereas the DOAs of interferers were 50°/80° and 50°/100°. We generated 36 test signals under the open window condition, where only IRs measured at the center of seats were used. For the closed window condition, we generated 180 test signals using IRs measured at all the positions. All the audio files were downsampled to 8 kHz. We computed STFT using a hamming window whose length was set at 128 ms and the window shift was 32 ms. The details of the dataset are shown in Table I.

B. Methods and evaluation criteria

We compared GCAV-IVA using fixed and data-adaptive constraints with AuxIVA [17], the DS beamformer, the MVDR beamformer, and the maximum signal-to-noise (maxSNR) beamformer [23]. These methods can be categorized into three classes of how much prior information they need namely, BSS, methods that need the DOA of the target, and methods that need training samples containing the target active period or/and interferer active period.

For AuxIVA and GCAV-IVA, we run 50 iterations to estimate the separation matrices, which were initialized with identity matrices. Since AuxIVA is a BSS method, the order of output channels is arbitrary. We evaluated outputs from all the channels and took the best score as a result. The hyperparameters c and λ for GCAV-IVA are shown in Table II. We used the same target active period and interferer active period as prior information for MVDR, maxSNR beamformers, and data-adaptive GCAV-IVA. The intervals of these periods were 5 seconds long.

TABLE III: Average SDR, SIR, and SAR [dB] over test dataset achieved by different method.

Method		SDR	SIR	SAR
BSS	AuxIVA	13.43	20.86	14.45
fixed constraints	DS	0.60	0.79	**17.00**
	GCAV-IVA(1)	6.15	10.56	9.17
data-adaptive constraints	MVDR	10.97	**24.02**	11.23
	maxSNR	13.43	20.63	14.45
	GCAV-IVA(2)	**14.25**	21.70	15.19

TABLE IV: Average SDR [dB] under conditions where car window was open or closed

Method		closed	open
BSS	AuxIVA	12.95	12.35
fixed constraints	DS	0.52	0.58
	GCAV-IVA(1)	7.35	5.84
data-adaptive constraints	MVDR	9.82	11.55
	maxSNR	12.92	11.89
	GCAV-IVA(2)	**13.98**	**12.93**

TABLE V: Average SDR [dB] under conditions where the interferer was at rear left of right seat.

method		rear left	rear right
BSS	AuxIVA	13.63	13.23
fixed constraints	DS	0.68	0.52
	GCAV-IVA(1)	7.12	5.18
data-adaptive constraints	MVDR	13.15	8.79
	maxSNR	13.75	13.10
	GCAV-IVA(2)	**14.64**	**13.85**

The speech enhancement performance was evaluated using the signal-to-distortion ratio (SDR), source-to-interferences ratio (SIR), and sources-to-artifacts ratio (SAR) [24].

C. Results and discussion

Table III shows the average SDR, SIR, and SAR scores over the entire test dataset. We found that GCAV-IVA outperformed baseline beamforming methods in each category, and data-adaptive GCAV-IVA achieved the highest SDR score of 14 dB among all the methods. For methods where the DOA of the target was known, GCAV-IVA achieved significant improvement in terms of both SDR and SIR whereas it was inferior to DS beamformer in terms of SAR. On the basis of beam pattern shown in Fig. 2, one possible reason for these results is that three microphones were insufficient to form a sharp beam to suppress the interferers, although they led to fewer artifacts in signals. For data-adaptive methods, although MVDR achieved the highest SIR score, which was about 2.5 dB higher than that achieved by GCAV-IVA, and the performance difference between maxSNR and GCAV-IVA was insignificant, GCAV-IVA was noteworthy in that only the target-active period was needed as prior information. In contrast, MVDR and maxSNR beamformers needed both target and interferer-active periods. These results confirmed the effectiveness of GCAV-IVA functioning for an in-car speech enhancement system.

We then show the results under more specific conditions. Table IV shows the results of conditions where the car window was open or closed. To make a fair comparison, the average SDR under the closed condition was computed over test signals using IRs recorded in the center position. The difference between the results under the two conditions was slight when using AuxIVA and DS beamformer, whereas it was significant when using other methods. Both maxSNR beamformer and GCAV-IVA achieved higher scores under the closed window condition whereas the MVDR beamformer performance was good under the open window condition. Table V shows the results under conditions where the interferer was set at the rear left or right seat. Similarly, AuxIVA and DS beamformer achieved comparable results under both conditions. All the other methods achieved better speech enhancement performance when the interferers were set at the passenger and rear left seats. Compared with the rear right seat having close DOA to the driver seat, the passenger and rear left seats are far from the driver seat in terms of DOA, which is considered to be a simpler condition for performing spatial filtering.

IV. Conclusion

In this study, we evaluated the effectiveness of GCAV-IVA as an in-car speech enhancement system in realistic environment. GCAV-IVA is a directional speech enhancement method that combines IVA with beamforming-based linear constraints. We applied two approaches to obtain the steering vector, which is necessary to conduct the constraints in GCAV-IVA. We compared the speech enhancement performance of GCAV-IVA with those of conventional beamforming methods and AuxIVA using in-car impulse responses measured with a triangle microphone array. The results revealed that GCAV-IVA outperformed all the baseline methods.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 19H04131 and Strategic Core Technology Advancement Program (Supporting Industry Program).

References

[1] I. Lecomte, M. Lever, J. Boudy and A. Tassy, “ Car noise processing for speech input, ” in *Proc. ICASSP*, 1989.

[2] O. Ichikawa, T. Fukuda, and M. Nishimura, “ Local peak enhancement for in-car speech recognition in noisy environment, ” in *IEICE Transactions on Information and Systems.*, vol. 91, no. 3, pp. 635-639, 2008.

[3] J. T. Chien, and P. Y. Lai, “ Car speech enhancement using microphone array beamforming and post filters, ” in *the 9th Australian International Conference on Speech Science & Technology Melbourne*, pp. 568-572, 2002.

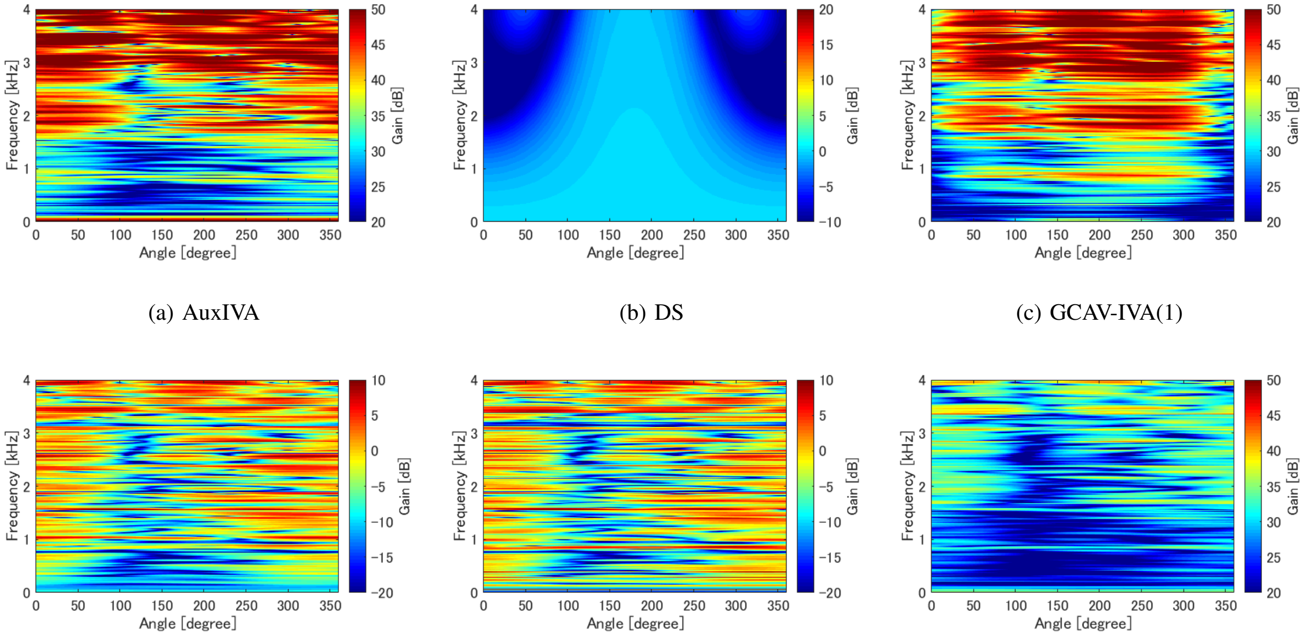
[4] T. Nakatani, M. Delcroix, and M. Fujimoto, “ Speech enhancement in a car using spatial and spectral models for speaker and noise, ” in *The 6th Biennial Workshop on Digital Signal Processing for In-Vehicle Systems* pp. 89-92, Sep., 2013.

[5] S. Boll, “ Suppression of acoustic noise in speech using spectral subtraction, ” in *IEEE Trans. Acoust., Speech and Signal Process*; pp. 113-120, 1979.

[6] Harry L., and Van Trees, “ Optimum array processing: Part IV of detection, estimation, and modulation theory, ” 2004, John Wiley & Sons.

[7] M. Wölfel, and J. McDonough, “ Distant speech recognition, ” 2009, John Wiley & Sons.

[8] L. C. Parra, and C. V. Alvino, “ Geometric source separation: Merging convolutive source separation with geometric beamforming, ” *IEEE Trans. SAP*, vol. 10, no. 6, pp. 352–362, 2002.



(a) AuxIVA

(b) DS

(c) GCAV-IVA(1)

(d) MVDR

(e) maxSNR

(f) GCAV-IVA(2)

Fig. 2: Example of beam patterns obtained by each method

[9] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, “ Blind source separation based on a fast-convergence algorithm combining ICA and beamforming, ” *IEEE Trans. ASLP*, vol. 14, no. 2, pp. 666–678, 2006.

[10] A. H. Khan, M. Taseska, and E. A. P. Habets, “ A geometrically constrained independent vector analysis algorithm for online source extraction, ” in *Proc. LVA/ICA*, pp. 396–403, 2015.

[11] L. Li, and K. Koishida, “ Geometrically constrained independent vector analysis for directional speech enhancement,” in *Proc. ICASSP*, pp. 846–850, 2020.

[12] T. Kim, T. Eltoft, and T.-W. Lee, “ Independent vector analysis: An extension of ICA to multivariate components, ” in *Proc. ICA*, pp. 165–172, 2006.

[13] A. Hiroe, “ Solution of permutation problem in frequency domain ICA using multivariate probability density functions,” in *Proc. ICA*, pp. 601–608, 2006.

[14] N. Ono and S. Miyabe, “ Auxiliary-function-based independent component analysis for super-gaussian sources, ” in *Proc. LVA/ICA*, pp. 165-172, Sep. 2010.

[15] Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo, “ Vectorwise coordinate descent algorithm for spatially regularized independent low-rank matrix analysis, ” in *Proc. ICASSP*, pp. 746–750, 2018.

[16] N. Makishima, Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo, “ Independent deeply learned matrix analysis with automatic selection of stable microphone-wise update and fast sourcewise update of demixing matrix, ” Signal Processing (Elsevier), vol.178, no.107753, 12 pages, 2021.

[17] N. Ono, “ Stable and fast update rules for independent vector analysis based on auxiliary function technique, ” in *Proc. WASPAA*, pp. 189–192, 2011.

[18] J. Bourgeois and W. Minker, Eds., “ Linearly constrained minimum variance beamforming, ” pp. 27–38, 2009.

[19] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, “ A consolidated perspective on multimicrophone speech enhancement and source separation,” *IEEE/ACM Trans. ASLP*, vol. 25, no. 4, pp. 692–730, 2017.

[20] D. R Hunter and K. Lange, “ A tutorial on MM algorithms, ” *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[21] O. L. Frost, “ An algorithm for linearly constrained adaptive array processing, ” *Proc. IEEE* vol. 60, no. 8, pp. 926–935, 1972.

[22] A. Kurematsu, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, and K. Shikano, “ATR Japanese speech database as a tool of speech recognition and synthesis, ” Speech Communication, vol. 9, pp. 357–363, 1990

[23] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. Saruwatari, “ Blind speech separation in a meeting situation with maximum SNR beamformers, ” in *Proc. ICASSP* vol. 1, pp. I–41, 2007

[24] E. Vincent, R. Gribonval, and C. Fevotte, “ Performance measurement in blind audio source separation, ” *IEEE Trans. ASLP,* vol. 14, no. 4, pp. 1462–1469, 2006.